

Keyword and Phrase Extraction from Persian Texts: A Review of the literature

Atefeh Kalantari*

PhD Candidate; Knowledge and Information Science Department;
School of Education & Psychology; Shiraz University; Shiraz, Iran;
Academic Librarian; Nursing & Midwifery Faculty;
Qazvin University of Medical Sciences; Qazvin, Iran;
Email: atefeh_kalantari@shirazu.ac.ir

Abdolrasool Jowkar

PhD in Knowledge & Information Science; Professor; School
of Education & Psychology; Shiraz University; Shiraz, Iran;
Email: ajowkar2003@yahoo.com

Seyed Mostafa Fakhrahmad

PhD in Computer Engineering; Associate Professor; Department
of Computer Science and Engineering & IT; Shiraz University;
Shiraz, Iran Email: fakhrahmad@shirazu.ac.ir

Javad Abbaspour

PhD in Knowledge & Information Science; Assistant Professor;
School of Education & Psychology; Shiraz University; Shiraz, Iran;
Email: javad.abbaspour@gmail.com

Hajar Sotudeh

PhD in Knowledge & Information Science; Associate Professor;
School of Education & Psychology; Shiraz University; Shiraz, Iran;
Email: sotudeh@shirazu.ac.ir

Massoud Mortazavi

PhD Candidate; Knowledge & Information Science Department;
School of Education & Psychology; Ferdowsi University
of Mashhad; Mashhad, Iran; Research & Development Specialist;
Pars Azarakhsh Co.; Tehran, Iran;
Email: mortazavi.massoud@mail.um.ac.ir

Amir Javadi

PhD in Medical Informatics; Assistant Professor; Social Medicine
Department; Qazvin University of Medical Sciences; Qazvin, Iran;
Email: javadi_a@yahoo.com

Zahra Pourbahman

PhD Candidate in Computer Engineering; AmirKabir University
of Technology; Tehran, Iran Email: pourbahman@aut.ac.ir

* Corresponding Author

Abstract: Keyword and phrase extraction is a prerequisite of many natural language processing tasks. However, a review on the related Persian and English literature showed that a few studies have already been done on how to extract keywords and phrases from Persian texts. Thus, aiming to shed light on the research status of Keyword and phrase extraction from Persian texts, the present study reviews the Persian and English publications which have assessed their research ideas over Persian texts. We also focus on each of the studies to challenge their methodologies, implementations and evaluation methods and measures.

To our knowledge, a total number of 14 Persian and 6 English papers exist which have worked on the extraction of Persian keywords and phrases. Investigating on the papers revealed that they were mostly based on statistical and linguistic information. A majority of the papers suffered from the lack of either appropriate methodologies or lucid explanation of their research ideas. They generally used non-standard datasets and vague or problematic metrics to evaluate the experimental systems. Generally speaking, except 3 papers that appropriately reported their proposed methods, the other papers lacked reproducibility and generalizability. Hence, their results cannot be confidently used as a benchmark in evaluating future works, and their proposed ideas cannot be employed in developing applications for extraction of key words and phrases from Persian texts.

Keywords: Extraction, Key Words, Key Phrases, Natural Language Processing, Persian Language, Review

Iranian Research Institute
for Information Science and Technology
(IranDoc)

ISSN 2251-8223

eISSN 2251-8231

Indexed by SCOPUS, ISC, & LISTA

Vol. 36 | No. 2 | pp. 563-592

Winter 2021



استخراج کلمات و عبارات کلیدی از متون فارسی

(مروری بر پژوهش‌های صورت گرفته)

عاطفه کلاتری

دانشجوی دکتری علم اطلاعات و دانش‌شناسی (بازیابی اطلاعات)؛ دانشگاه شیراز؛ شیراز، ایران؛ مسئول کتابخانه دانشکده پرستاری و مامایی؛ دانشگاه علوم پزشکی قزوین؛ قزوین، ایران؛
atefeh_kalantari@shirazu.ac.ir

عبدالرسول جوکار

دکتری علم اطلاعات و دانش‌شناسی؛ استاد؛ بخش علم اطلاعات و دانش‌شناسی؛ دانشکده روان‌شناسی و علوم تربیتی؛ دانشگاه شیراز؛ شیراز، ایران؛
ajowkar2003@yahoo.com

سیدمصطفی فخر احمد

دکتری مهندسی کامپیوتر؛ دانشیار؛ بخش مهندسی و علوم کامپیوتر و فناوری اطلاعات؛ دانشگاه شیراز؛ شیراز، ایران
fakhrahmad@shirazu.ac.ir

جواد عباس‌پور

دکتری علم اطلاعات و دانش‌شناسی؛ استادیار؛ بخش علم اطلاعات و دانش‌شناسی؛ دانشکده روان‌شناسی و علوم تربیتی؛ دانشگاه شیراز؛ شیراز، ایران؛
javad.abbaspour@gmail.com

هاجر ستوده

دکتری علم اطلاعات و دانش‌شناسی؛ دانشیار؛ بخش علم اطلاعات و دانش‌شناسی؛ دانشکده روان‌شناسی و علوم تربیتی؛ دانشگاه شیراز؛ شیراز، ایران؛
sotudeh@shirazu.ac.ir

مسعود مرتضوی نصرآباد

دانشجوی دکتری علم اطلاعات و دانش‌شناسی (بازیابی اطلاعات)؛ دانشگاه فردوسی مشهد؛ مشهد، ایران؛ کارشناس بخش تست و توسعه؛ شرکت پارس آذرخش؛ تهران، ایران
mortazavi.massoud@mail.um.ac.ir

امیر جوادی

دکتری انفورماتیک پزشکی؛ استادیار گروه پزشکی اجتماعی؛ دانشکده پزشکی؛ دانشگاه علوم پزشکی قزوین؛ قزوین، ایران
javadi_a@yahoo.com

زهرا پوربهمن

دانشجوی دکتری مهندسی کامپیوتر (نرم‌افزار)؛ دانشکده مهندسی کامپیوتر؛ دانشگاه صنعتی امیرکبیر؛ تهران، ایران؛
pourbahman@aut.ac.ir



مقاله برای اصلاح به مدت ۳۷ روز نزد پدیدآوران بوده است.

پدیش: ۱۳۹۹/۰۳/۲۵

دریافت: ۱۳۹۸/۰۷/۱۸

چکیده: استخراج کلمات/عبارات کلیدی متن، پیش‌نیاز بسیاری دیگر از وظایف حوزه پردازش زبان طبیعی است. اما بررسی متون فارسی و انگلیسی این حوزه نشان می‌دهد که تلاش‌های انگشت‌شماری برای استخراج کلمات/عبارات کلیدی از متون فارسی صورت گرفته است. از این‌رو، این مقاله با هدف تعیین موقعیت کنونی پردازش زبان طبیعی فارسی، و به‌طور خاص، استخراج کلمات/عبارات کلیدی از متون فارسی به مرور خلاصه‌ای از مقالات فارسی و انگلیسی منتشرشده در این حوزه که

نشریه علمی | رتبه بین‌المللی
پژوهشگاه علوم و فناوری اطلاعات ایران
(ایرانداک)

شاپا (جایی) ۲۳۵۱-۸۲۳۳

شاپا (الکترونیکی) ۸۳۳۱-۲۳۵۱

نمایه در SCOPUS، ISI، و LISTA

jipm.irandoc.ac.ir

دوره ۳۶ | شماره ۲ | صص ۵۶۳-۵۹۲

زمستان ۱۳۹۹

